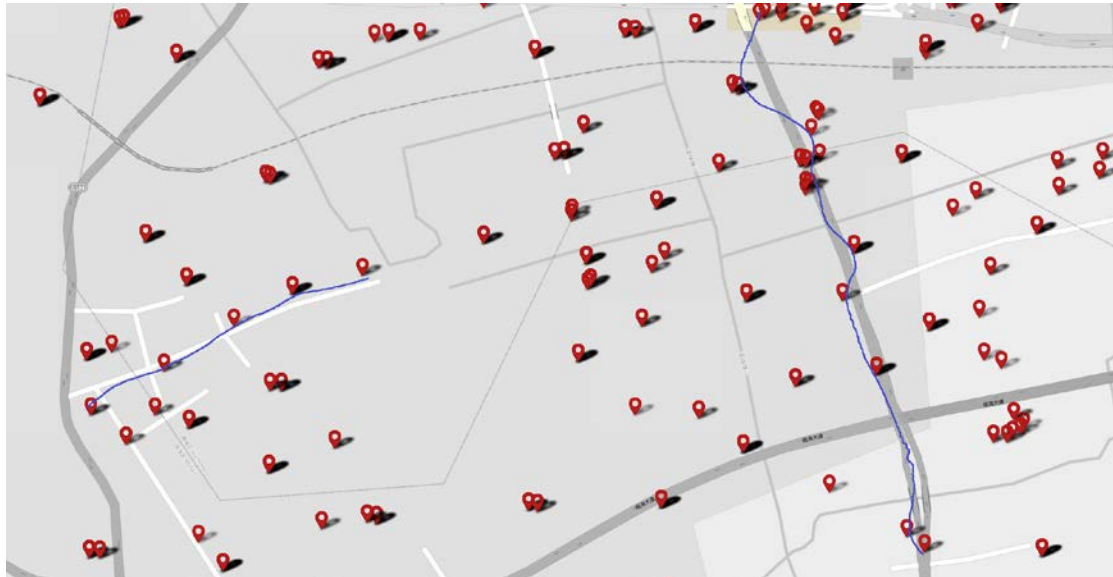


结合目前正在研究的出租车轨迹数据和基站人群移动数据，我试图将沿道路查询的查询模型进行扩展。

1、扩展 VAST 论文投 TVCG

轨迹数据可以看到比较明显的模式。从组内同学的实验结果看，移动数据的模式也比较明显，因为在城市中人群的主要交通工具是汽车。而且从基站位置的布局来看，很多基站也都是沿着道路设置的（见附图）。或许两种数据结合后，可以将我们先前的工作扩充后转投 TVCG。比如说，可以从道路上的人流量来确定基站位置设置是否合理。



如果直接将我们的查询模型应用到移动数据会存在以下问题：

- 1) 基站数据范围比较大，无法用道路网覆盖基站数据
初步考虑是将基站数据投影到道路网上，对于离道路较远的点删除。
 - 2) 基站数据时间间隔较大而且间隔时长不规则，而我们的索引要求间隔时长一致。
初步考虑索引以小时为单位（出租车数据是以 20 秒为单位）
- 另外，如果投 TVCG，可视化设计也需要做调整。这部分内容还在考虑中。

2、将道路查询模型扩展成一般化的时空查询，投数据库相关会议（定位 VLDB）

模型扩展的方向主要是面向城市移动数据。查阅数据库方面的论文，还没有看到有相关文献。对于这个模型最关键的是查询效率。提高查询效率需要考虑两方面的问题：

- 1) 在分布式系统上实现查询算法
- 2) 缩小时空查询范围，减少扫描的数据量。

我觉得 TripHash 索引仍然可以继续使用，因为这是目前局部时空查询最有效的方式之一。本周分析了先前工作中的查询算法，发现算法中最主要的问题是：

- 网格单元中数据不均衡，有的网格数据量非常大，有的可能非常小
对于单元格一天以内的数据点其实并不多，而时间拉长后数据量增长较快，以前的工作没有处理时间维度上的问题（只做简单排序）。初步考虑使用类似空间填充曲线的三维（空间+时间）编码。GeoHash 是对空间进行编码的一种索引方式，B⁺树是对运动对象未来位置编码的一种索引。
- 查询每条轨迹时，轨迹经过的第一个网格中的数据都需要遍历一次（最坏全部遍历，最好只遍历前 N 条），如果单元格中数据量大遍历开销较大